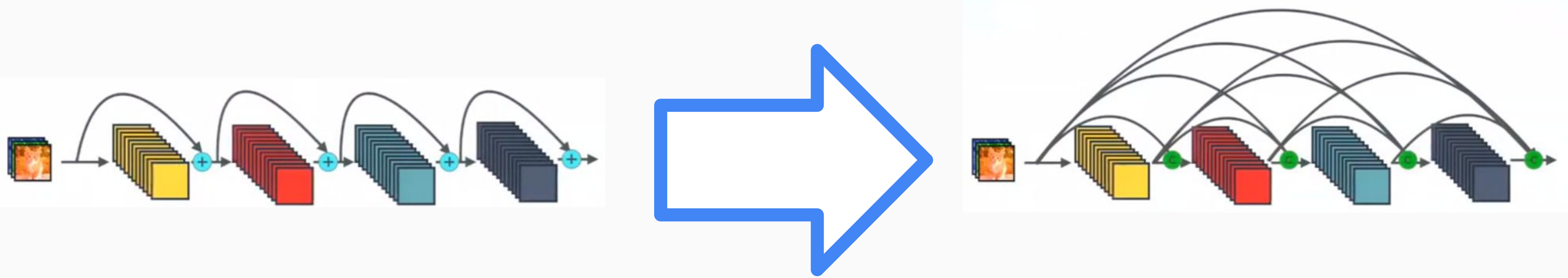
← Speaker: Rodrigo Benenson

Human-in-the-loop annotations

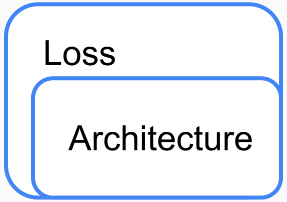
KEEP
CALM
AND
WRITE DOWN
QUESTIONS

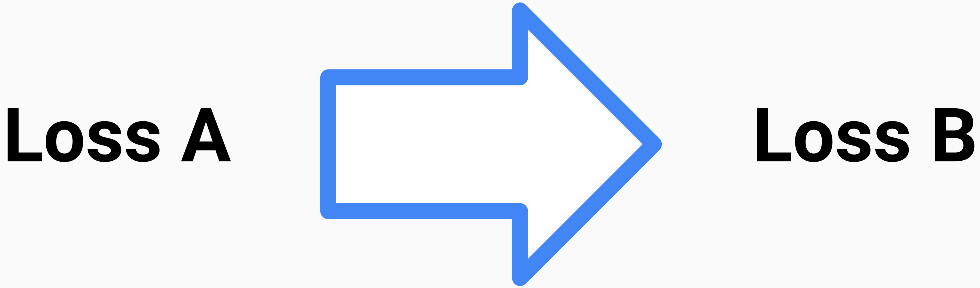
Typical machine learning paper focus on model training

Architecture

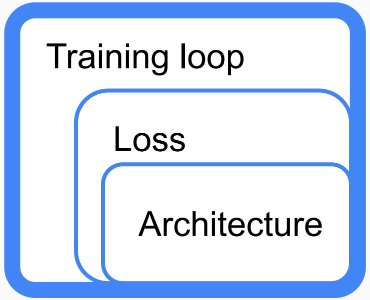


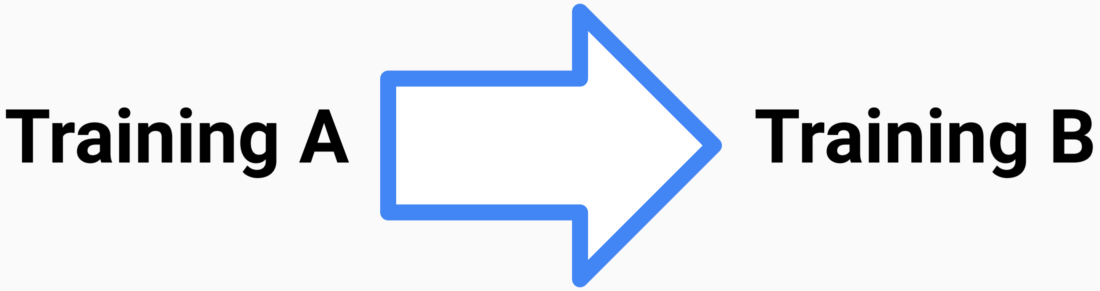
Typical machine learning paper focus on model training



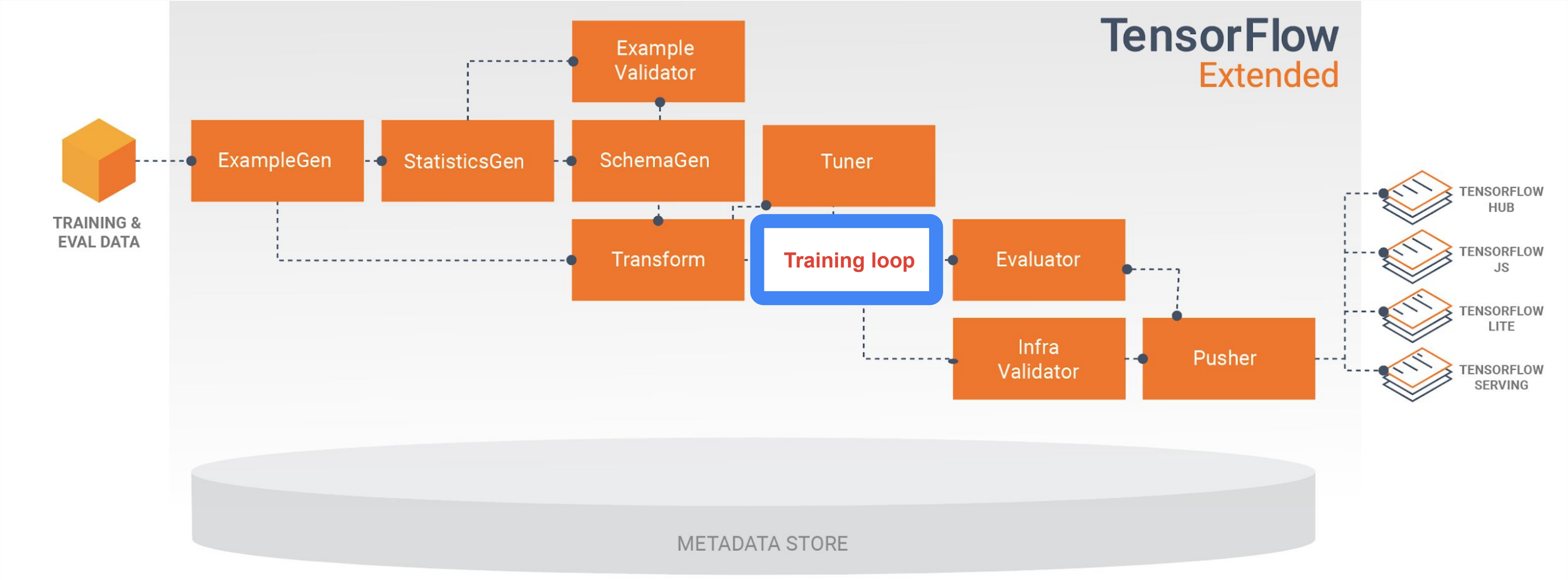


Typical machine learning paper focus on model training

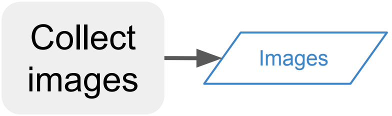




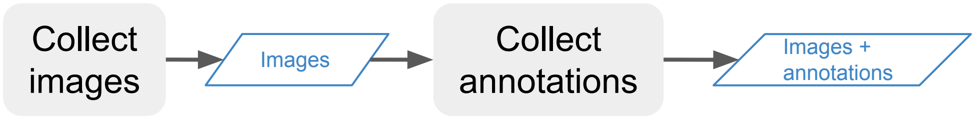
In practice machine learning is much more than data + model



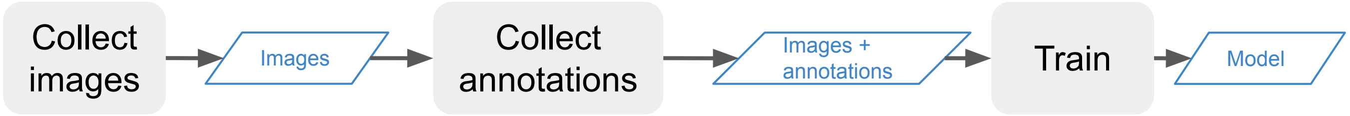
Data collection pipeline



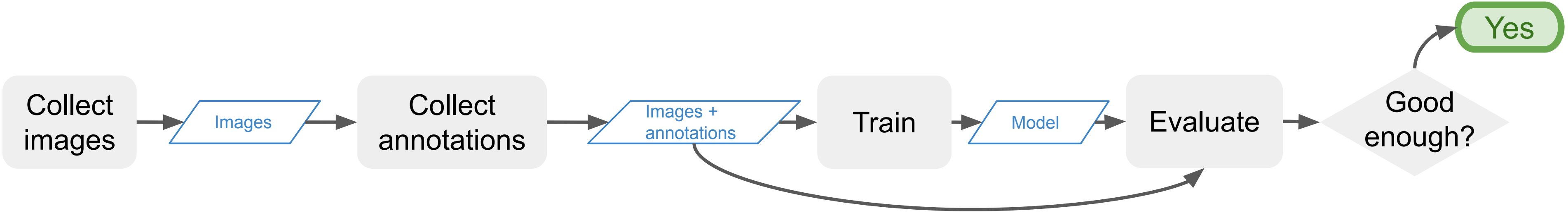
Data collection pipeline



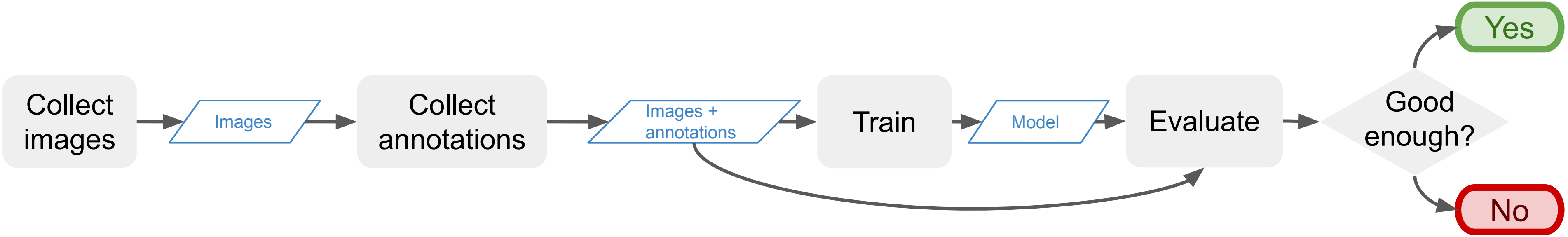
Data collection pipeline



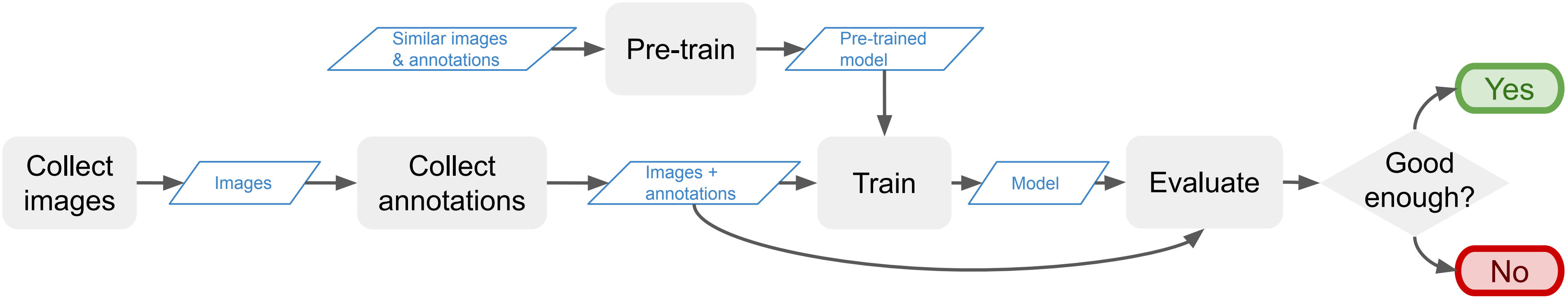
Data collection pipeline



Data collection pipeline

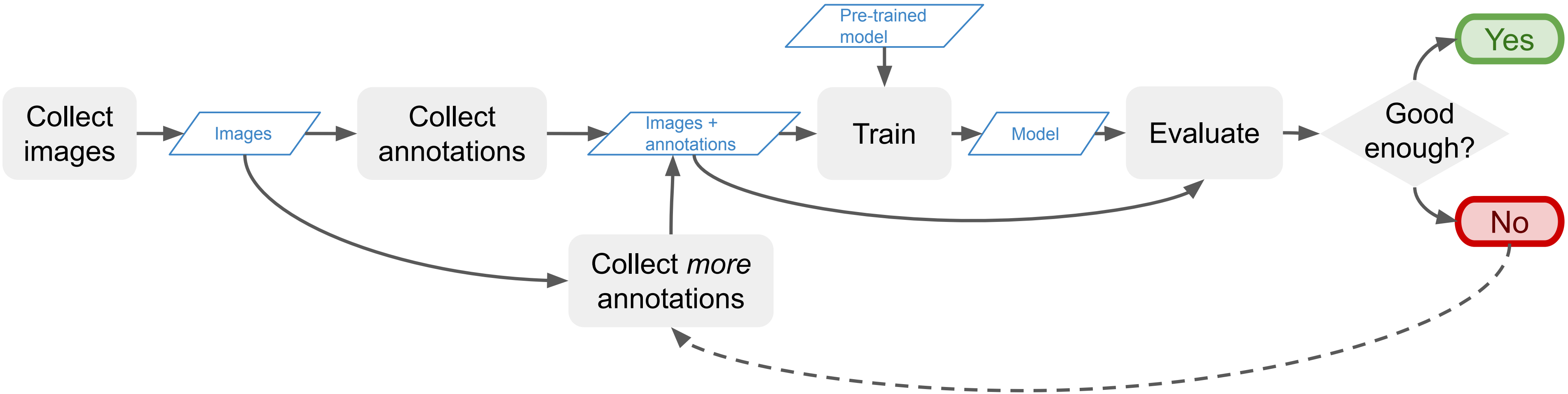


Data collection pipeline

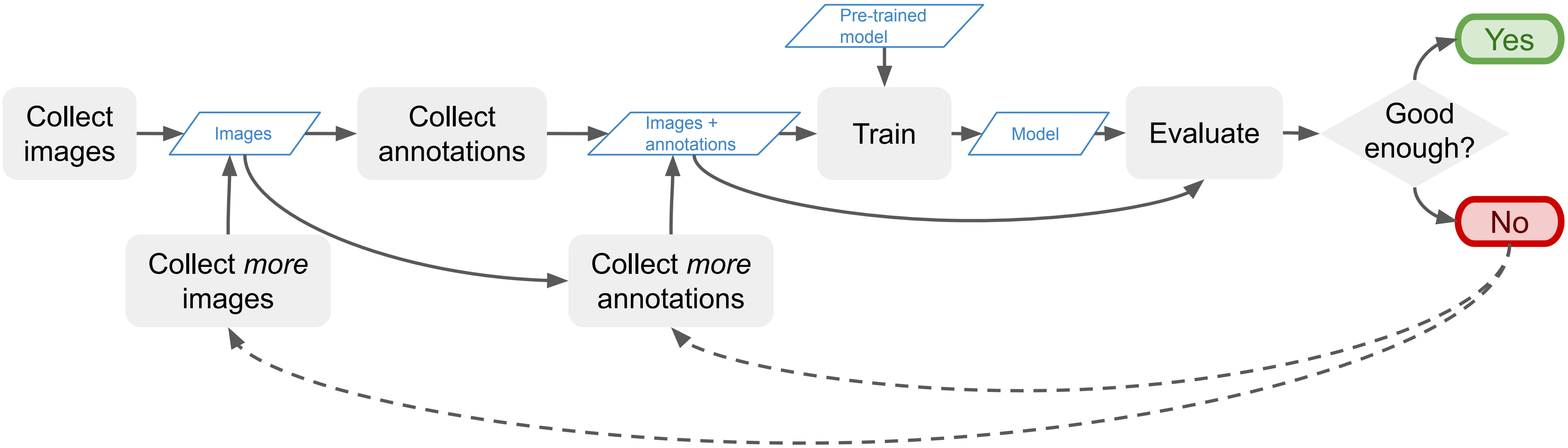


(In practice, transfer learning can be shockingly effective)

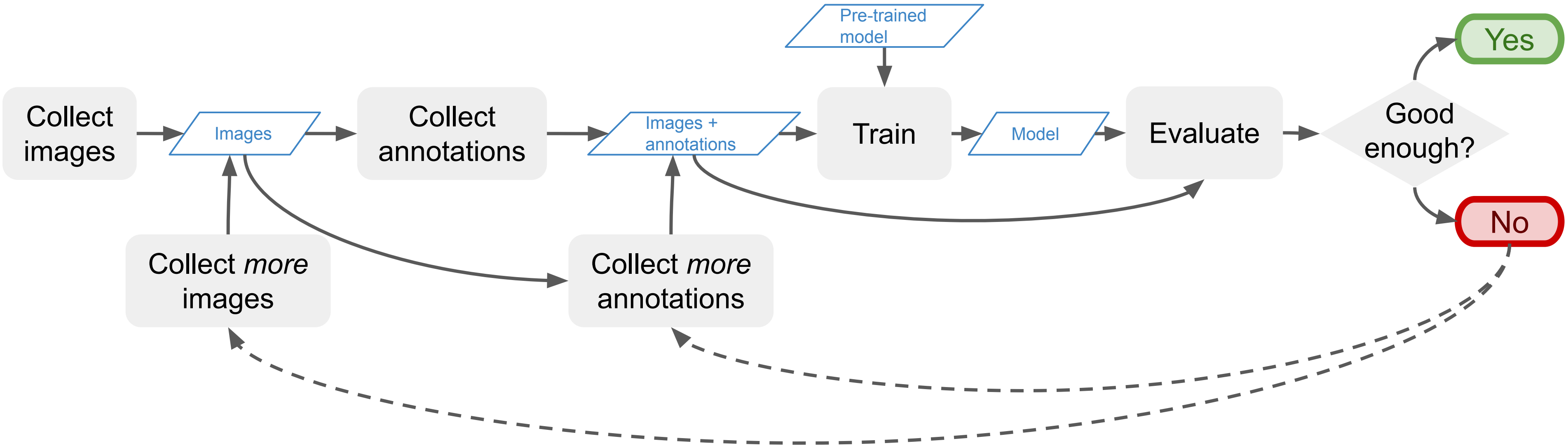
Data collection pipeline



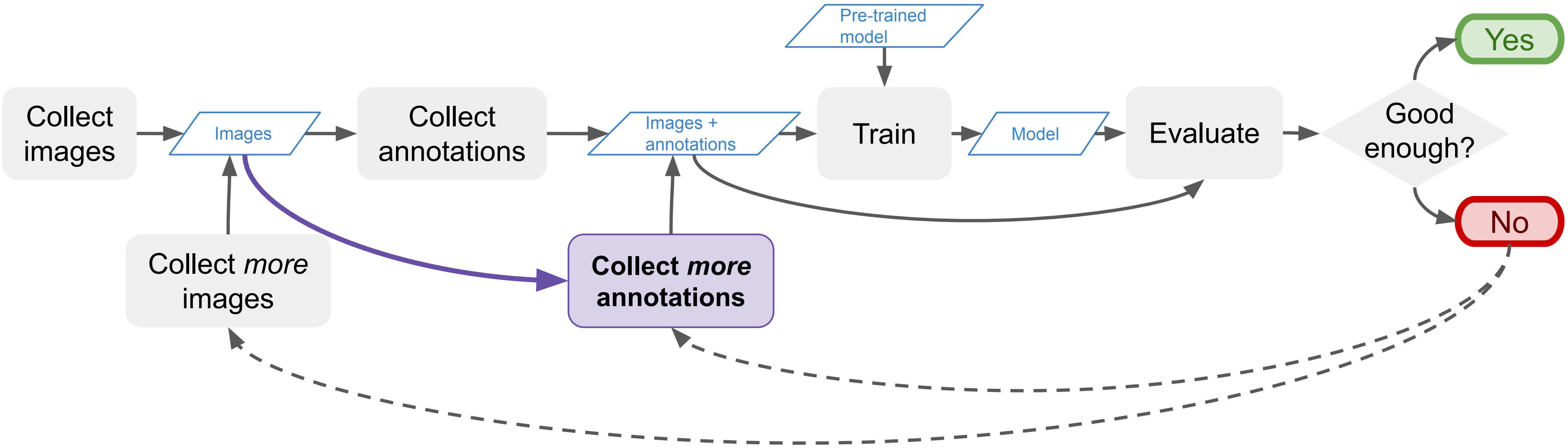
Data collection pipeline



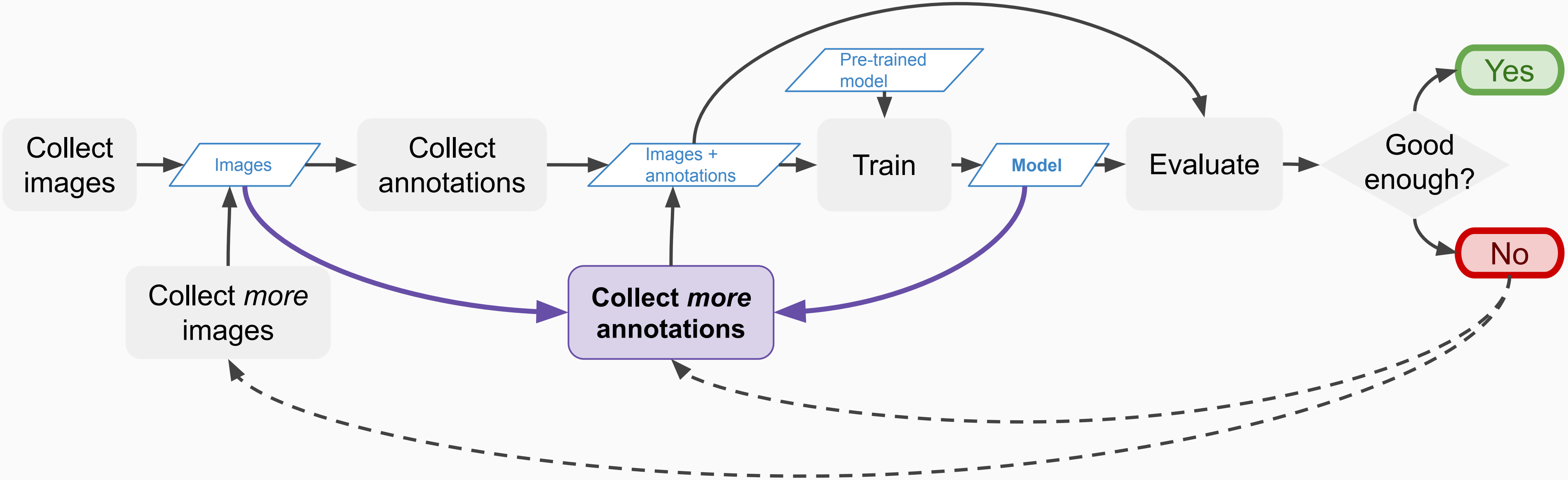
Data collection pipeline



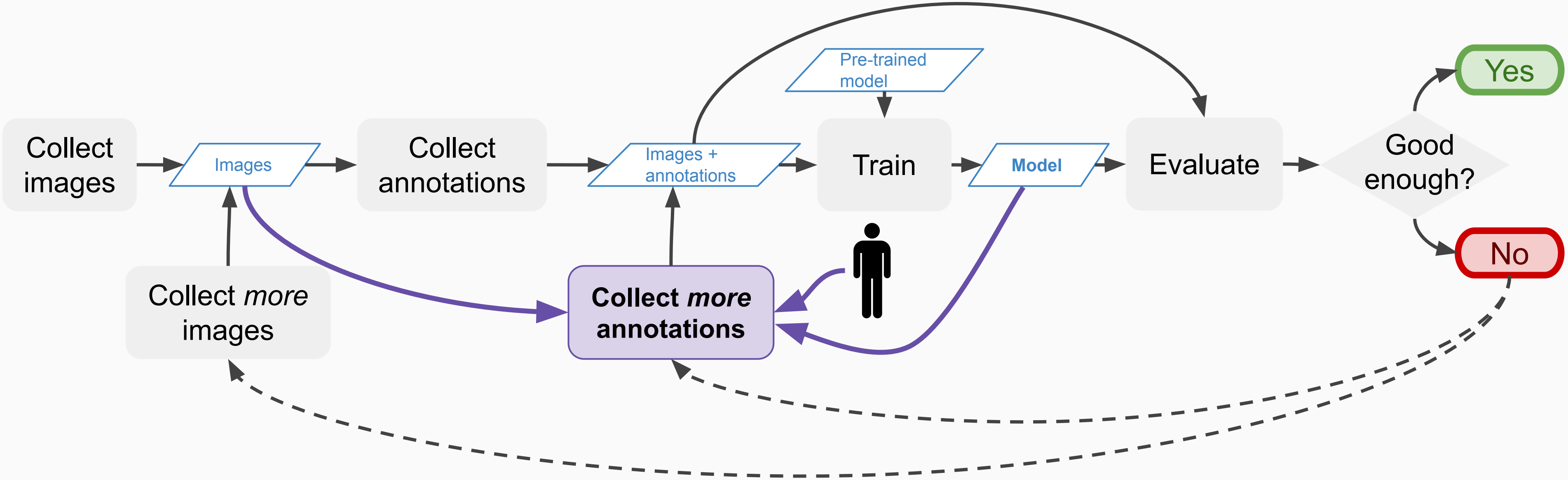
Data collection pipeline



The new annotations should be aware of the model



The new annotations should be aware of the model



The new annotations should be aware of the model

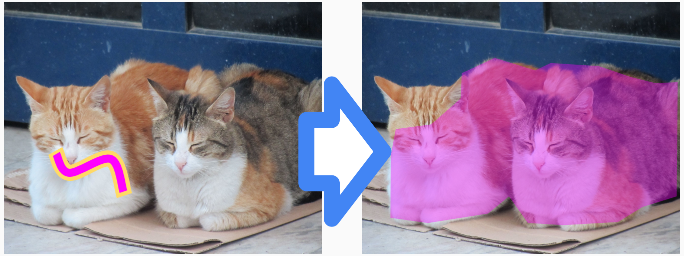
Before

The new annotations should be aware of the model

Before

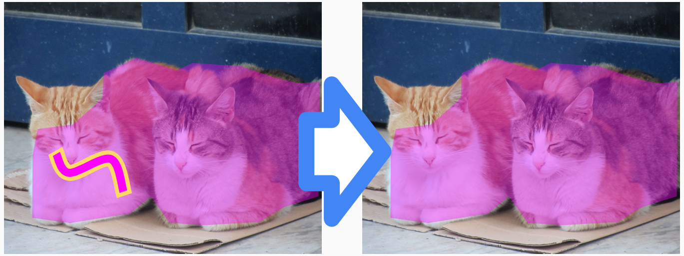
The new annotations should be aware of the model

Before After



The new annotations should be aware of the model

Before After



✗Redundant with what the model already knew.

The new annotations should be aware of the model

Before

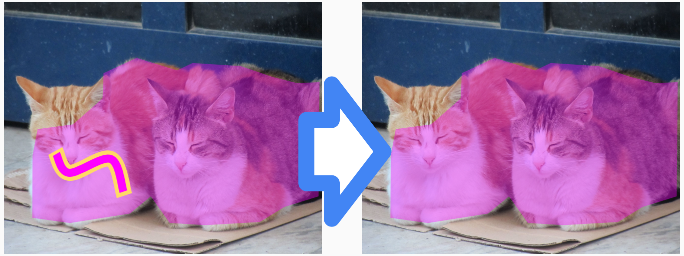
After

✗ Redundant with what the model already knew.

Before

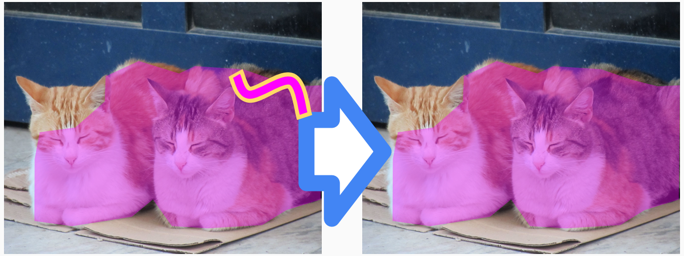
The new annotations should be aware of the model

Before After



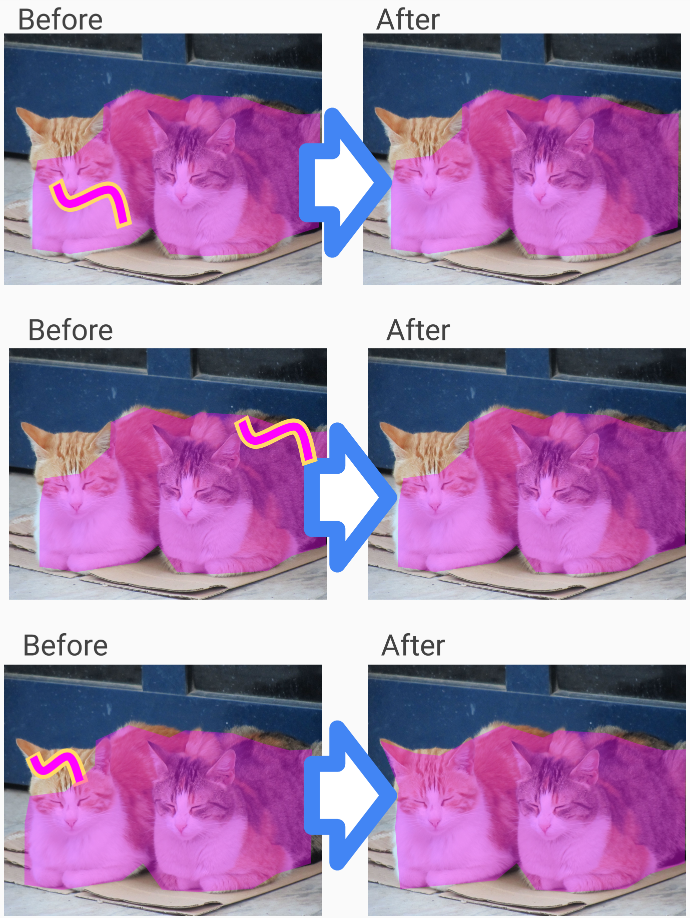
✗ Redundant with what the model already knew.

Before After



✗ Too hard for the model to learn.

The new annotations should be aware of the model

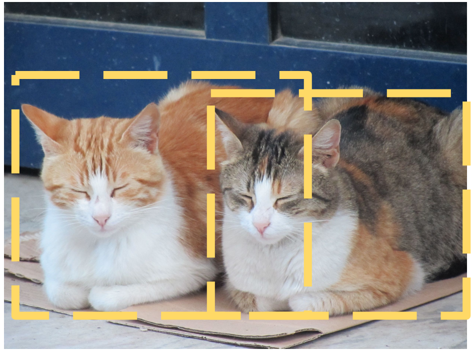


✗ Redundant with what the model already knew.

✗ Too hard for the model to learn.

✓ **Informative** and **learnable** annotation.

Detection



Which boxes to add?

Semantic labeling

Which pixels to add?

Which image areas should be annotated ?

(aka active learning)

Which annotation
will lead to an improved model ?

Which annotation
will lead to an improved model ?

⇒ Hard problem

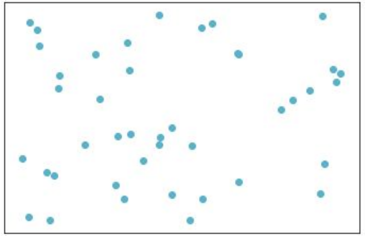
Which annotation
will lead to an improved model ?

⇒ Hard problem

Heuristics

Annotation selection heuristics

- **Uniform**
- Score bands
- High entropy
- Ensemble disagreements
- Self-consistency



Uniform: accept one's ignorance.

Pros:

- As simple as it gets.
- Reasonable strategy to bootstrap annotations.

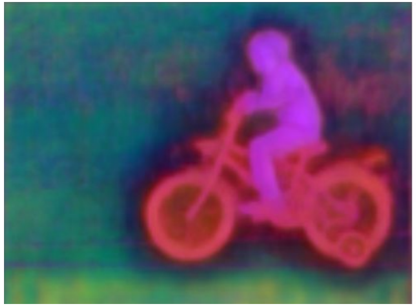
Cons:

- If model is reasonably good,
  high portion of redundant annotations.
- If class distribution is skewed,
  will under-represent some classes.

Variant (if image-level labels):
uniform annotations,
but only across the bottom-N worst classes.

Annotation selection heuristics

- Uniform
- **Score bands**
- High entropy
- Ensemble disagreements
- Self-consistency



Score band: focus on areas with score ∈ [a, b].

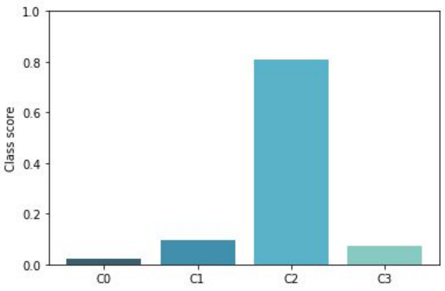
E.g. score ∈ [0.4, 0.6], score ∈ [0.8, 0.9].

Pros:
- Simple to implement.
- Can easily target ambiguous regions.
- Can aim for class-balanced sampling.

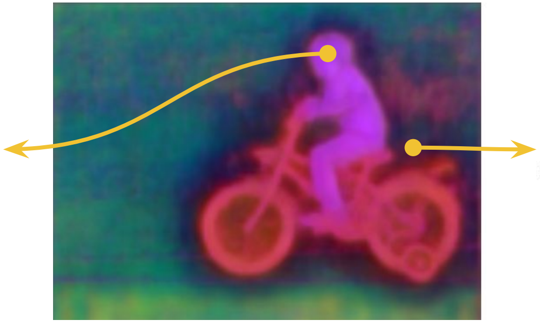
Cons:
- Empirically not very effective.

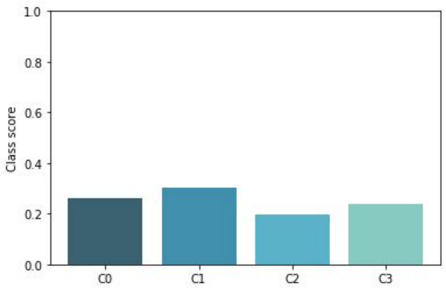
Annotation selection heuristics

- Uniform
- Score bands
- **High entropy**
- Ensemble disagreements
- Self-consistency



Low confusion





High confusion

High entropy: focus on areas of model confusion.

$$H(x) = -\sum_k p_k \log(p_k)$$

Pros:

- Simple to understand.
- Annotated samples guaranteed to provide training loss.
- Empirically hard to beat.

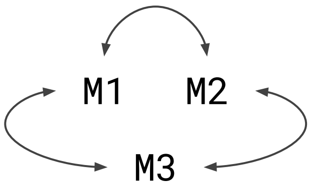
Cons:

- Does not include a notion of sample diversity.

Annotation selection heuristics

- Uniform
- Score bands
- High entropy
- **Ensemble disagreements**
- Self-consistency

Ensemble disagreements:
focus where N models disagree.



Disagreement measured by l2-norm,
Jensen-Shannon divergence, vote entropy, etc.

Pros:
- Better estimation of model uncertainty.
- Provides better results than single model.

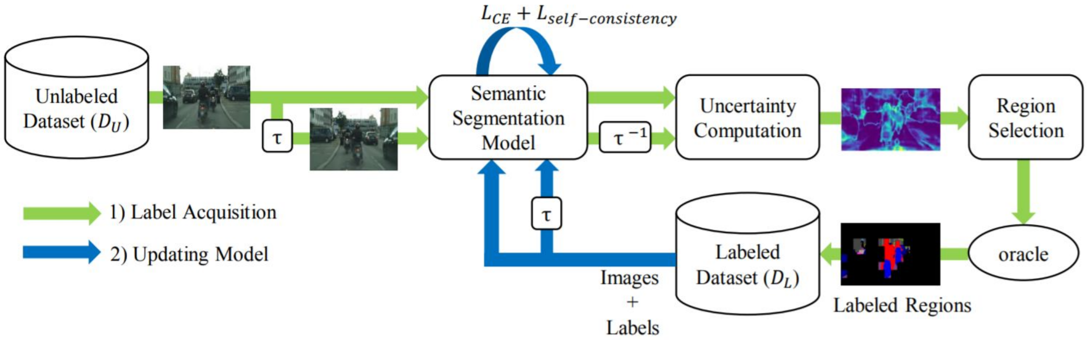
Cons:
- Requires training multiple models.
  (Ensemble can be approximated via dropout)

(Ensemble average can also be used as a single stronger model, and use high-entropy)

[Dagan & Engelson ICML 95]

Annotation selection heuristics

- Uniform
- Score bands
- High entropy
- Ensemble disagreements
- **Self-consistency**



Self-consistency:
focus where equivariance is not respected.

Pros:
- Simple to understand.
- Can (should) be combined with the previous heuristics.

Cons:
- (Requires hand-crafting the test-time augmentation).

[Golestaneh & Kitani arxiv 2020]

Opinion: active learning is a field where most ideas do not work.

(most ideas work a little, sometimes)

If in doubt: ensemble model + entropy + self-consistency.

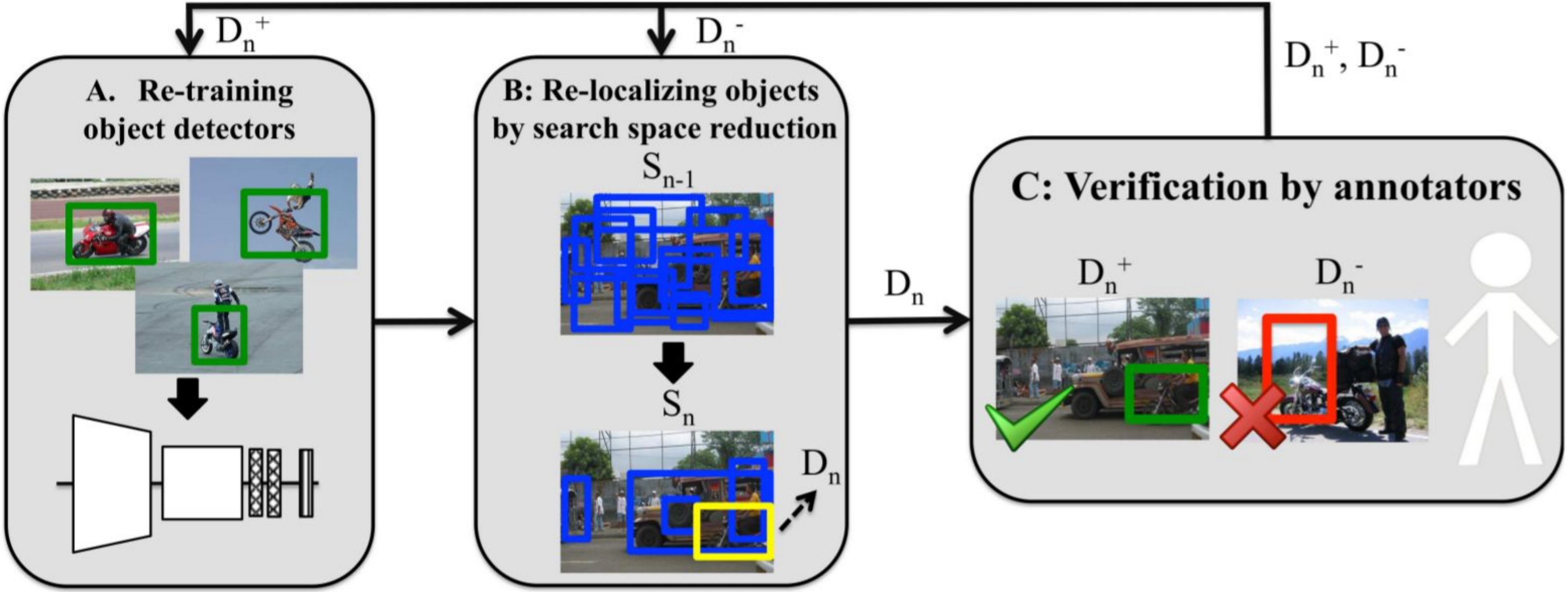
[BALD, LAL, RALIS, CEREALS]

Collecting bounding boxes

(without drawing any box)

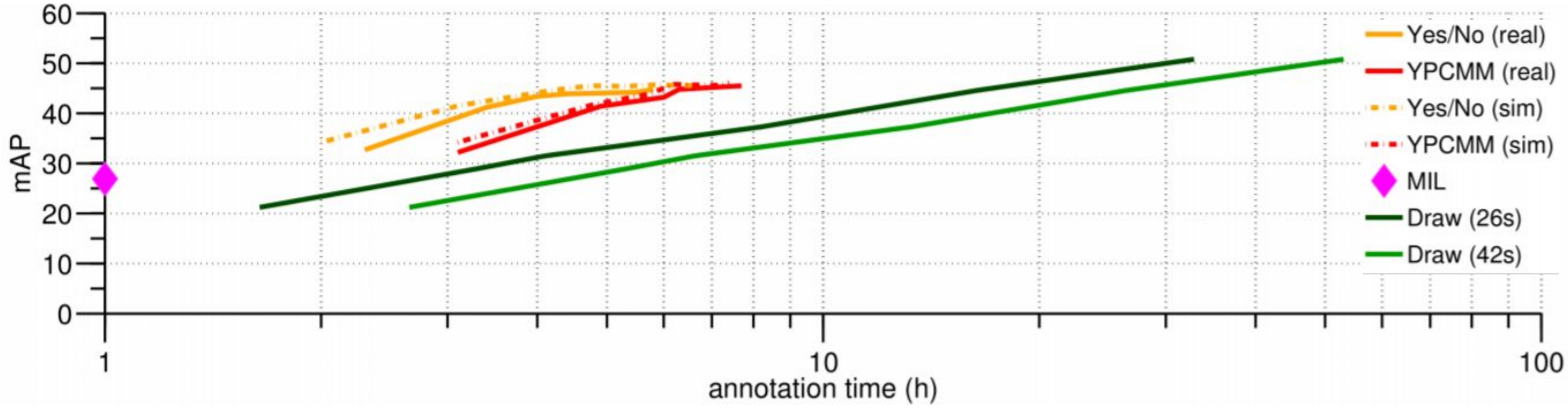
The annotator verifies boxes instead of drawing them

(yes/no or yes/part/container/mixed/missed)



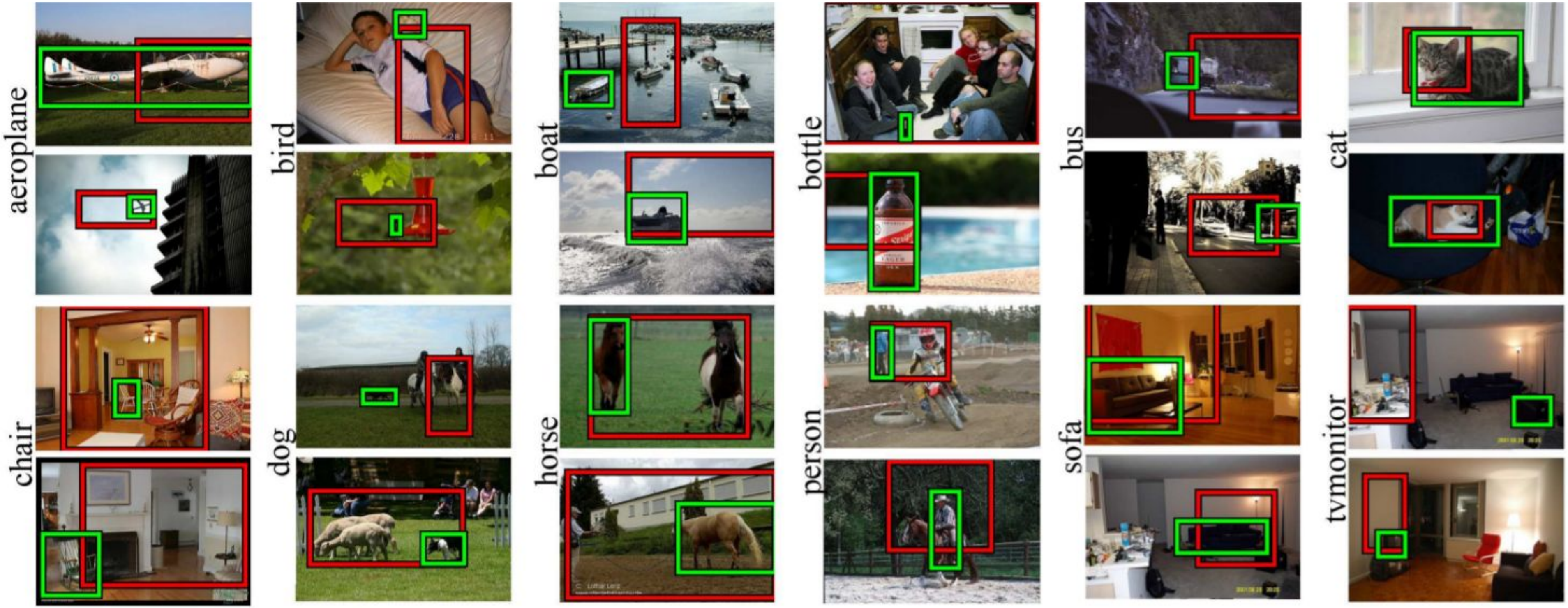
[Papadopoulos et al. CVPR 2016]

Better model when limited human time budget



Pascal VOC 2007 object detection evaluation.

[Papadopoulos et al. CVPR 2016]



Red: weakly supervised bounding boxes (from image-level labels).
Green: boxes after collecting verifications.

[Papadopoulos et al. CVPR 2016,
see also Kao ACCV 2018, Pardo et al. arxiv 2019]

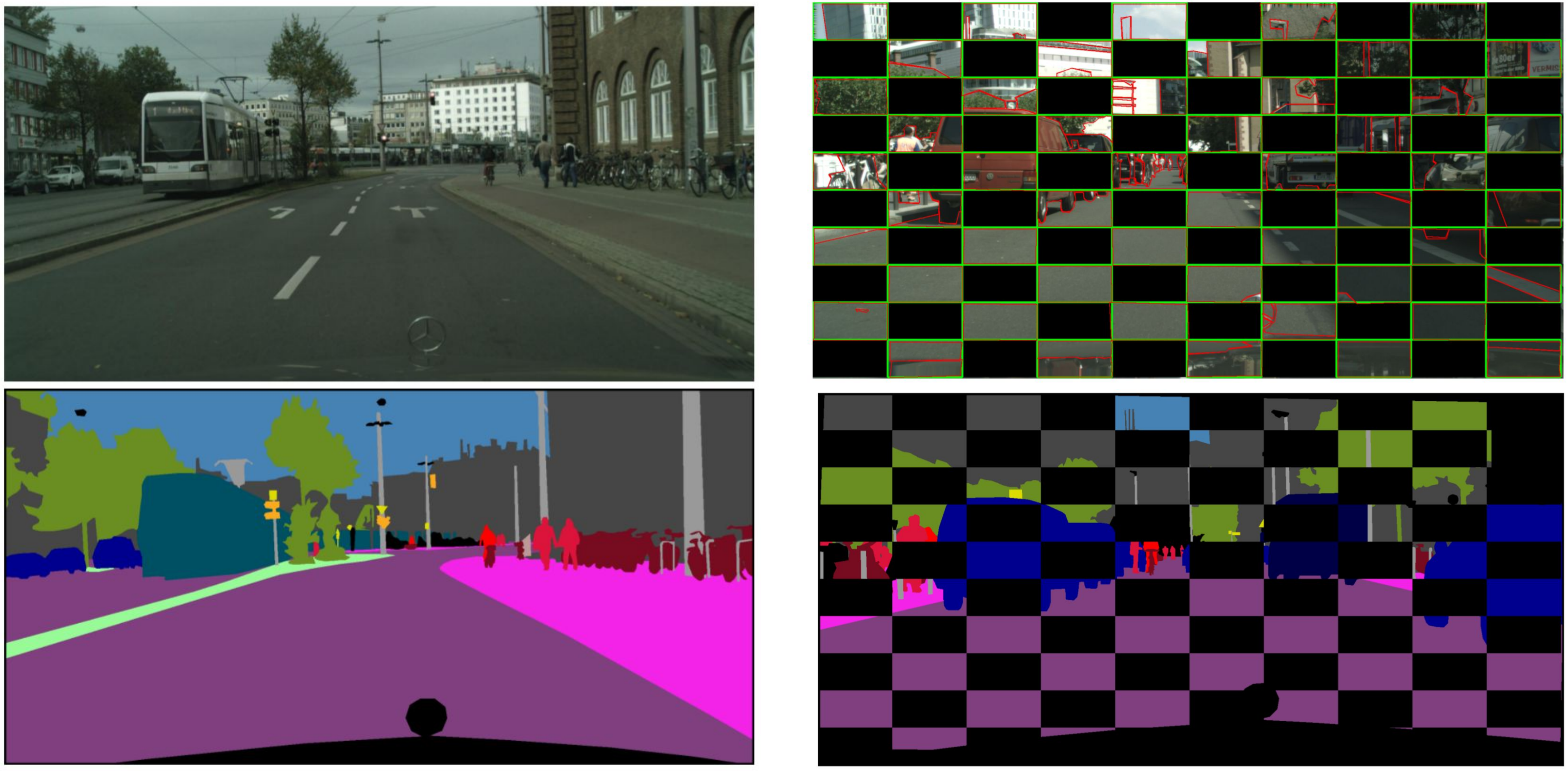
Collecting segmentations

(guiding the drawing hand)

Segmentation annotations do not need to be complete

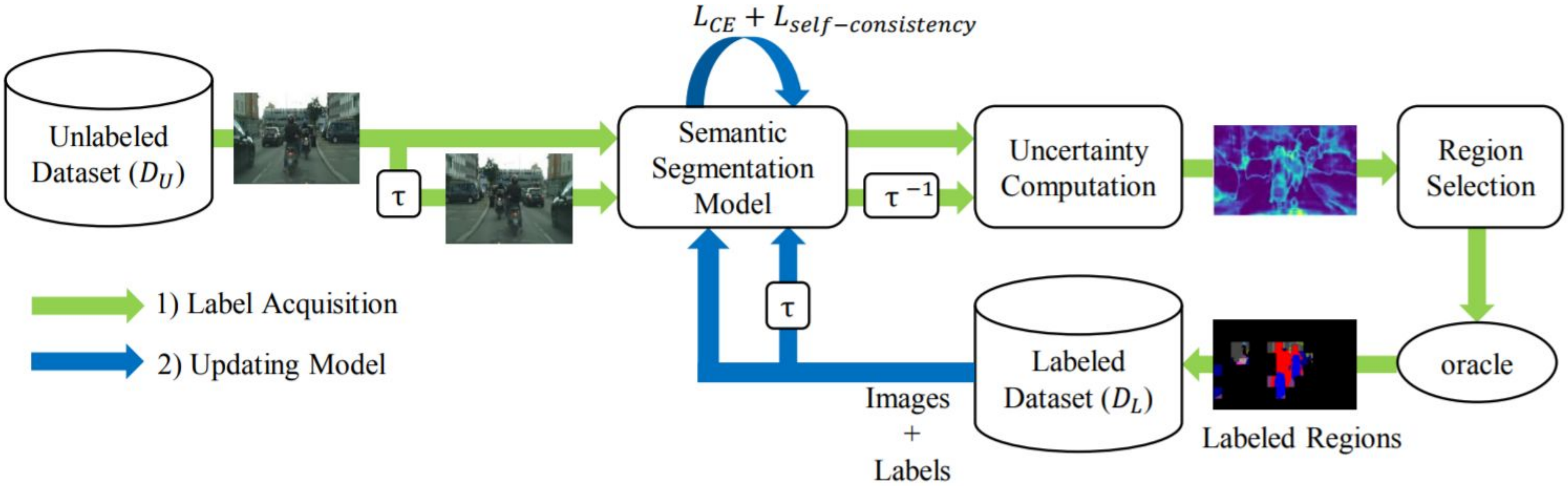
[Lin et al. ICCV 2019]

Segmentation annotations do not need to be complete



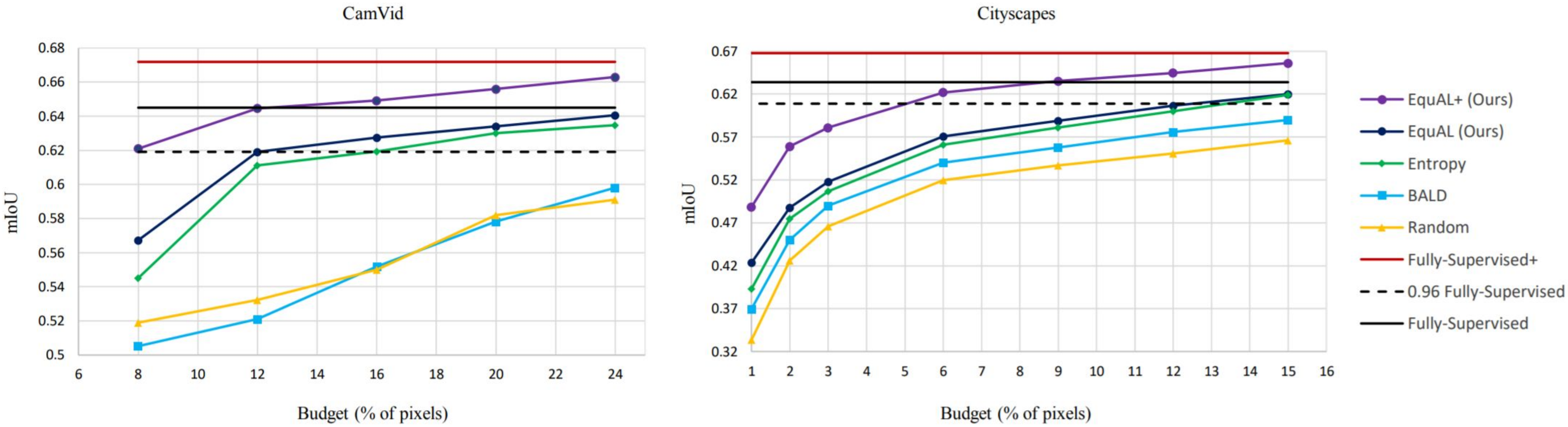
[Lin et al. ICCV 2019]

Segmentation blocks can be machine-selected



[Golestaneh & Kitani arxiv 2020]

Segmentation blocks can be machine-selected



[Golestaneh & Kitani arxiv 2020]

Collecting segmentations

(guiding the clicking hand)

Focusing user actions to errors

(per instance, masks corrective clicks)

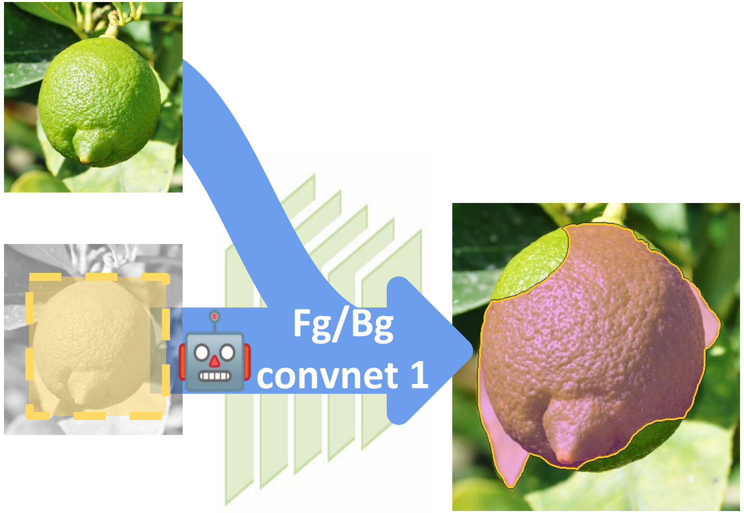
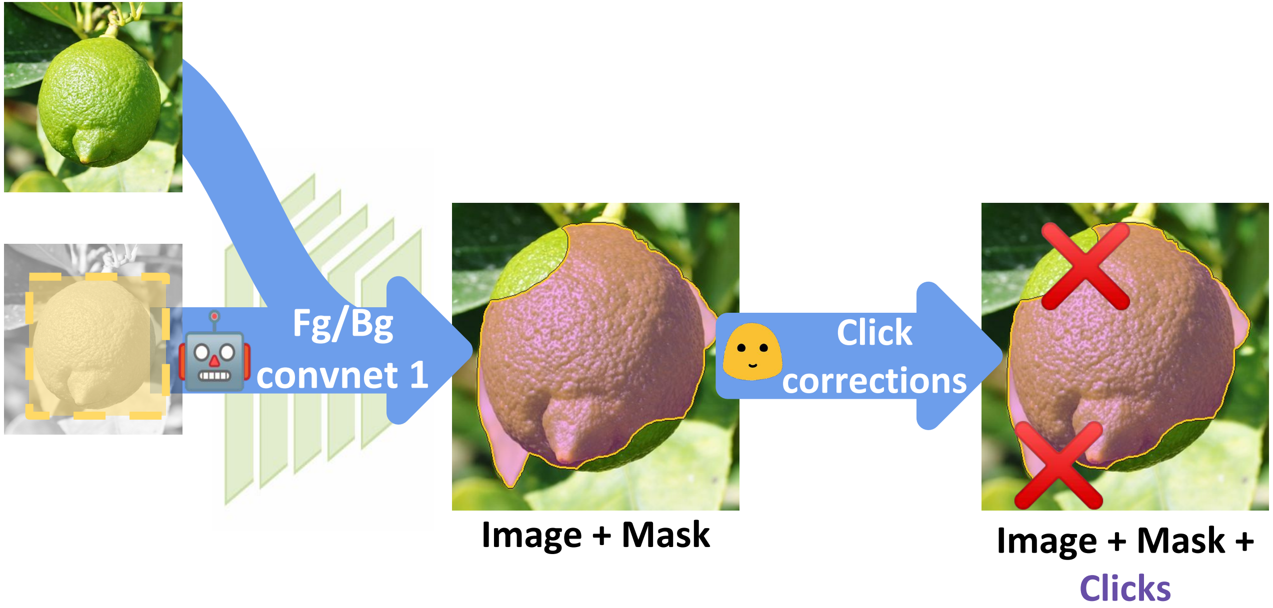


Image + Mask

[Benenson et al. CVPR 2019]

Focusing user actions to errors

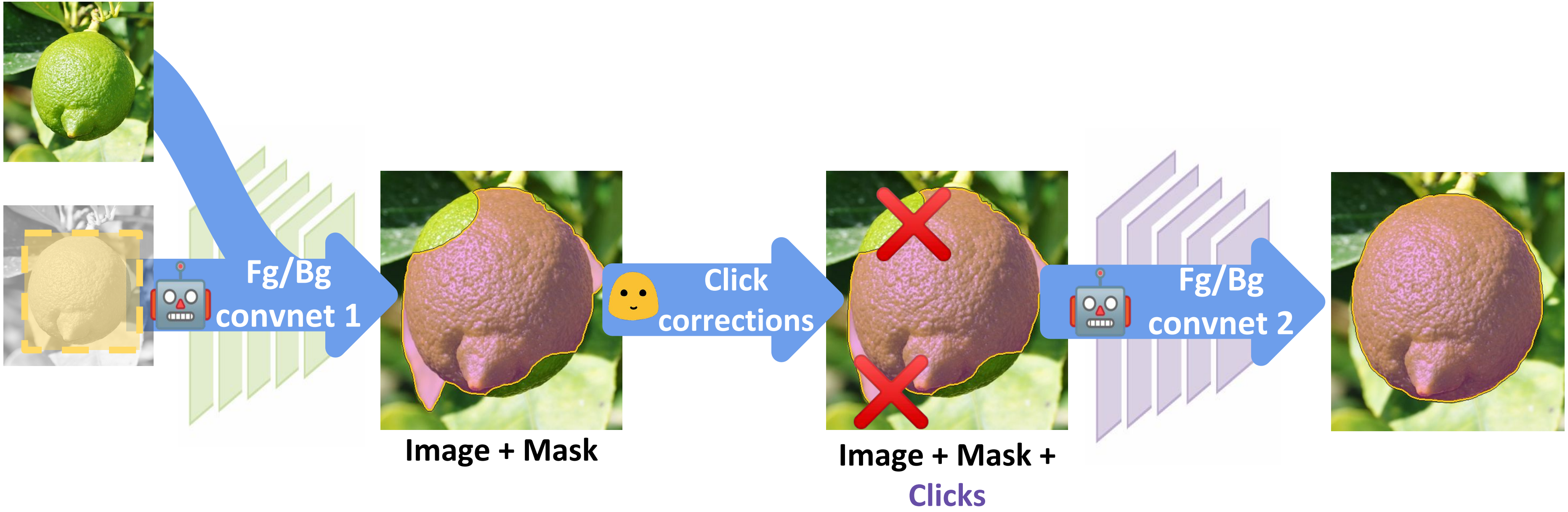
(per instance, masks corrective clicks)



[Benenson et al. CVPR 2019]

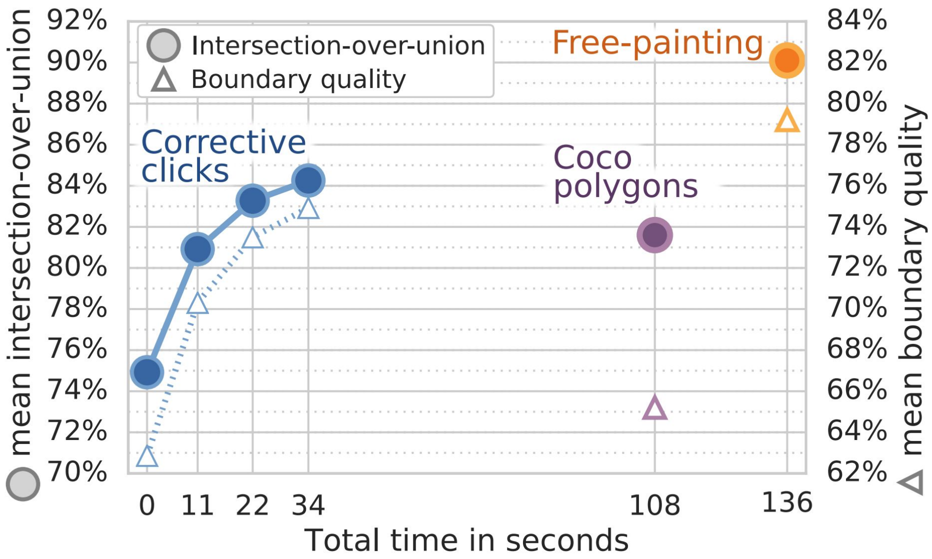
Focusing user actions to errors

(per instance, masks corrective clicks)



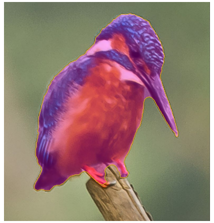
[Benenson et al. CVPR 2019]

Focusing user actions to errors
(per instance, masks corrective clicks)



- Quality > COCO polygons
- ~3x faster annotation time
- 2.5M instances masks

https://g.co/dataset/open-images

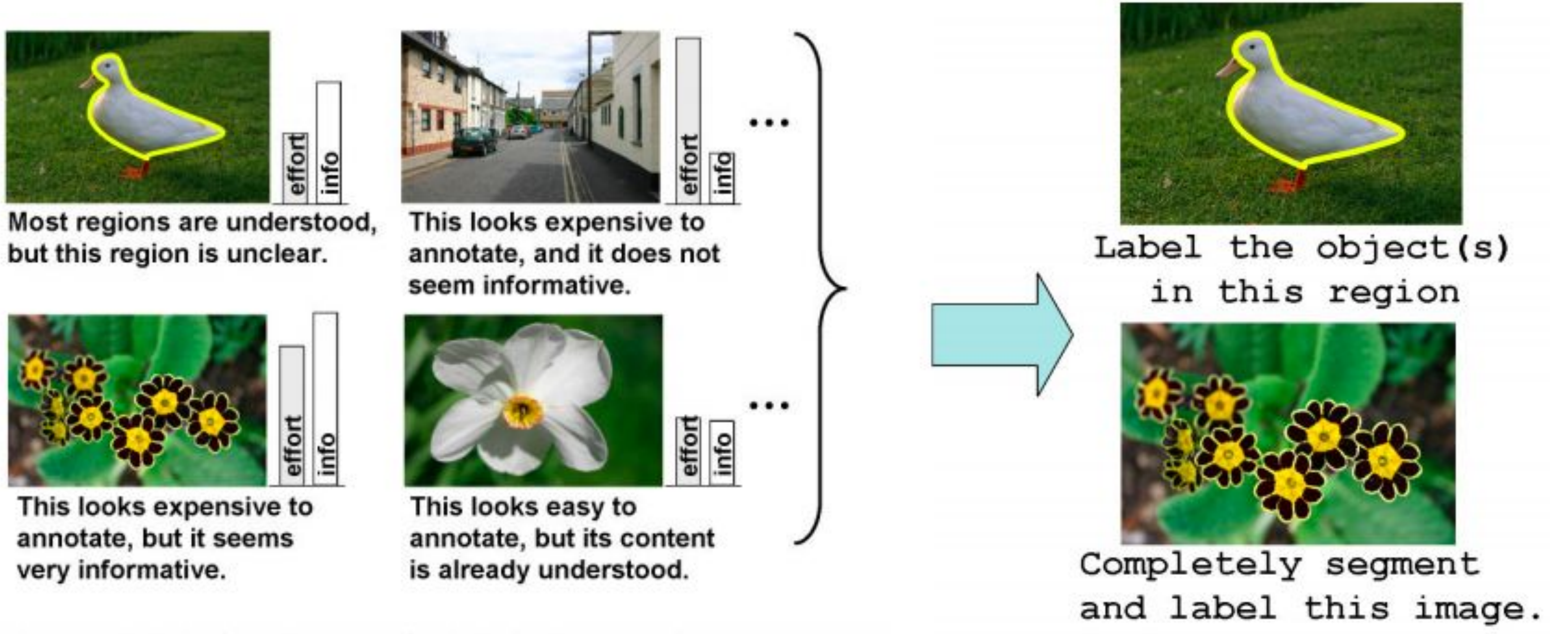



[Benenson et al. CVPR 2019, Kontogianni et al. ECCV 2020]

Annotation dialogs

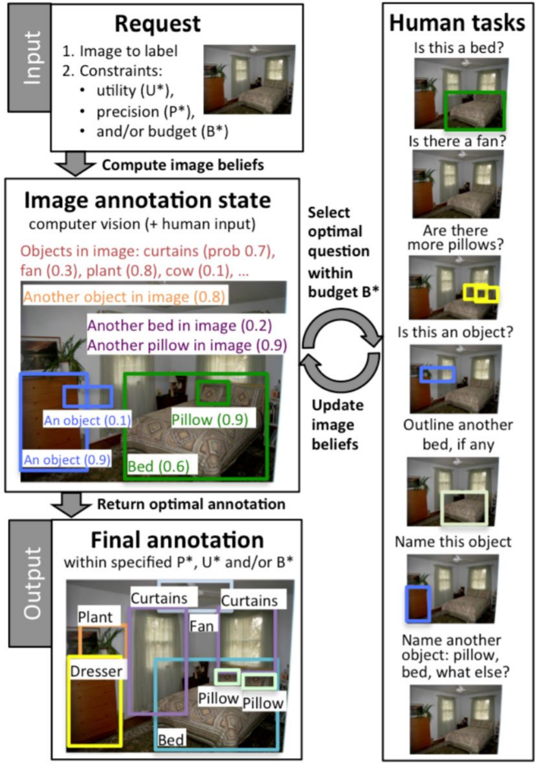
(where the machine ask)

The best strategy covers different annotation types, the machine asks what it needs.



[Vijayanarasimhan and Grauman CVPR 09, Russakovsky et al. CVPR 15, Konyushkova et al. CVPR 18]

The best strategy covers different annotation types, the machine asks what it needs.



[Vijayanarasimhan and Grauman CVPR 09, Russakovsky et al. CVPR 15, Konyushkova et al. CVPR 18]

Takeaways:

- For large scale annotation campaigns,
  hybrid annotations enable better use of human time.
- Strong annotations can be partial, and focused.
- For active learning component, keep it simple.
- Do not underestimate the power of transfer learning.
- There is a large design space for Human-Machine collaboration.

?